VIDEO ANALYSIS AND CONTENT EXTRACTION

TERRENCE ADAMS
DEPARTMENT OF DEFENSE
FORT MEADE, MD 20755-6514

Abstract– We describe the major objectives of the ARDA-VACE research and development program, and its impact on analysis of formal and informal meetings. This program is entering its second phase as it begins to handle a broad range of video related problems including, object detection and tracking, recognition, video enhancement, compression, video mining, multi-modal analysis, and video event representation and detection. We give a bird's eye view of the components of the program, including evaluations, and describe the exciting new technological advances and collaboration that occurs.

I. INTRODUCTION

The *Video Analysis and Content Extraction Program (VACE)* is one of several research and development programs run within *ARDA (Advanced Research and Development Activity*). It is currently entering its second phase, as it addresses multiple levels of video understanding including core, fusion and event-oriented tasks. There are fifteen primary projects covered by research centers from around the nation, working in five different video domains, including Foreign Broadcast News, Formal and Informal Meetings, Ground Reconnaissance, Surveillance and UAV Motion Imagery. Here we discuss the research objectives in these areas, and outline progress for *VACE* Phase II.

II. VACE OBJECTIVES

In the area of speech research, meeting room scenarios bring new challenges over telephone conversation and broadcast news data. Meetings may have many participants with several different speakers, some communicating simultaneously. In many cases, the collection of data through microphones and cameras is not a primary concern for those using the technology, and this makes content-based analysis difficult. There are many different organizational types of meetings, with many meetings being loosely formed. With these challenges, the information extracted from video and audio may be incomplete, and require higher level inferences to draw important conclusions.

In the first phase of *VACE*, research focused on the most relevant core level objectives including detection, tracking and extraction of text, people, faces and vehicles. In the second phase, research continues in these areas, as well as including video enhancement, compression, video stabilization, object recognition, video OCR, gesture analysis, query by example, event representation and detection. Underlying the video research is a study of the best pattern recognition, computational and machine learning techniques for handling these problems. Important advances may occur in these foundational areas which support the ultimate video objectives. In particular, recently, we have seen increased prominence for areas such as video ontologies and representation languages, protocols, graphical models, hidden Markov models, learning algorithms, frequency transformations as well as other advanced mathematical approaches. These areas are all important when seeking to intelligently decompose a problem into its constituent parts, and apply the best-known practices at each level. Video is a complicated media type, and the power and sophistication of human sensory perception is often underestimated for workers outside the speech and computer vision fields. Today, we see researchers continuing to find room for significant advancement on these problems. Yet, we continue to learn the important properties for video content analysis, and new great ways for representing these properties.

For the meeting room scenario, interest first centers on object detection such as faces and text, and then recognition of these objects. Higher level objectives include tracking of people, heads, hands, torsos and other objects, as well as multi-modal analysis with speech data. Eventually, the goal is to accurately make inferences about the type of meeting such as a group or board meeting, lecture type, military style or some other format. What happened during the meeting? What decisions were made?

Who are the major players, or opinion leaders? Where did the meeting take place, and what is the time frame for the meeting, and any actions resulting from the meeting? Determine the "who, what, where and when" of a meeting. This is an exciting challenge, most notably in cases when the data is noisy, and collection is not designed for in-depth post-meeting analysis using automatic computer vision algorithms.

III. VACE PROGRESS

Several sub-programs are established to ensure the success of *VACE* phase II. Evaluations are planned and carried out on component-level video objectives, which show importance across multiple domains. Great detail and care goes into designing and implementing protocols and metrics for content extraction, and for crafting consistent annotation guidelines for evaluators to follow. For example, in the case of face detection, there needs to be clear guidelines on what constitutes a face, and what measures should be used to locate and extract a face. Do we want to find faces, which appear in photographs in the background of a video scene? Should a drawing or painting of a face be labeled as a face for evaluation purposes? There is a subjective nature to these tasks, and ultimately the users of these technologies determine its functionality.

Since much of the research in phase II is unique and does not cut across several domains, we have instituted exploratory evaluations. This allows the researcher greater input in the design and implementation of these evaluations to correctly characterize the tasks. These evaluations are drawing an interest in face recognition, video enhancement and stabilization, as well as de-noising and video OCR.

Besides participation in evaluations, all researchers must demonstrate progress from initial concept models toward product implementation. What ideas will flourish and become attractive solutions for customers? We encourage progress along these lines, with the formulation of a video testbed. This highlights the importance of good programming practice and a sound strategy for research and development. This will ease the eventual transfer of technology into the hands of end-users.

IV. CONCLUSION

VACE is an exciting research and development program tackling some of the most challenging problems in video content analysis. From object detection and tracking, through recognition, to event representation and detection, significant achievements are being gained. There are still many more interesting problems to study, and new areas to explore. With the rapid rise in the volumes of video and other multi-media data, there is an ever-growing need for analytic capabilities against this data. *VACE* continues to address these problems head-on, breaking new ground on a day-to-day basis.

Filename: icassp_vace_2004
Directory: H:\VACE
Template: C:\Program Files\Microsoft Office\Templates\NORMAL.DOT
Title: The Video Analysis and Content Extraction Program is one of several research and development programs run within ARDA (Advanced Research and Development Activity)
Subject:
Author: tmadam2
Keywords:
Comments:
Creation Date: 04/20/04 1:46 PM
Change Number: 7
Last Saved On: 04/27/04 11:46 AM
Last Saved By: tmadam2
Total Editing Time: 11 Minutes
Last Printed On: 04/27/04 11:46 AM
As of Last Complete Printing
Number of Pages: 2
Number of Words: 991 (approx.)
Number of Characters: 5,650 (approx.)